

Technical Memorandum
Lower Passaic River Restoration Project
Guidance for Revising the Sediment Quality Triad (SQT) Analysis

December 11, 2015
Norm Richardson, Battelle
Chuck Nace, USEPA

This memorandum summarizes United States Environmental Protection Agency (USEPA) guidance for revising the Sediment Quality Triad (SQT) analysis included in the draft Baseline Ecological Risk Assessment (BERA) for the Lower 17 Miles of the Passaic River (LPR) (Windward, 2014). USEPA's evaluation of the draft analysis determined that there were a number of technical flaws that had bearing on the conclusion reached that physical conditions rather than sediment chemistry was primarily responsible for the documented laboratory toxicity. In addition to improperly defining the reference condition, the draft analysis included bivariate correlational analysis to detect relationships between chemicals of interest (COI) and biological response (either single species toxicity testing measured in the laboratory or benthic community structure metrics). Although a number of COIs were found to be significantly correlated with one or more measures of biological response, in large part, these findings were eliminated following alpha-adjustment to control for family-level comparisons (i.e., Bonferroni applications). Although this technique is a routinely applied procedure to control for "false positives" when conducting "family-level" comparisons as in the draft BERA, its application can result in actual relationships being overlooked (and potentially resulting in sediment remediation not being conducted to address unacceptable ecological risks where warranted). The primary weakness of Bonferroni correction procedures is that analytical interpretations depend on the number of statistical tests performed (Perneger, 1998) with the likelihood of Type II errors and the risk of overlooking significant relationships increasing as more comparisons are included in the analysis. The SQT study design (Windward, 2009) lacks statistical power necessary to test significance for over 500 pairwise comparisons between analytes and biological response that were included in the draft BERA.

This issue was discussed during a conference call between USEPA and the CPG in 28 July and USEPA agreed to provide some guidance on the conduct of an alternative approach involving multivariate analytical (MVA) techniques. Attachment A provides an outline of such an approach prepared by Dr. John Kern and is included for the CPG's consideration. Although USEPA believes that multivariate approaches offer promise in elucidating underlying structure in complicated datasets such as the 17-mile LPR, it is also acknowledged that there are relatively few benthic community assessment applications in the literature. In addition, MVA is also subject to the same statistical power concerns raised with the bivariate correlational approach. The attached guidance is an approach that the CPG may wish to consider if it determines that a MVA approach seems promising. As agreed to previously, USEPA will provide comments on any specific proposal advanced by the CPG.

However, USEPA also believes that a properly conducted bivariate correlational analysis would be beneficial in the SQT for the LPR. Recognizing that the objective of the SQT process is to determine whether site-related chemical stressors appear to be an important component stressor in structuring the benthic community rather than to identify specific chemical stressors¹, USEPA requests that the bivariate analysis presented in the draft BERA be revised with the Bonferroni corrections eliminated. Eleven benthic endpoints² were evaluated in the draft BERA (9 or 10 depending on the area of interest) and combined with the 55 COIs resulted in Bonferroni-adjusted α 's ranging from 0.000091 to 0.00011, which represented an extremely high bar for demonstrating statistical significance. This degree of protection (against Type 1 errors) is unnecessary (Perneger, 1998) and the role of the analyst, who determines how many comparisons merit evaluation, in detecting significant effects is evident. Moreover, many of the COIs and endpoints are not independent in their effect or biological attribute being evaluated. Even though considered a pseudo-concern in this type of application (Perneger, 1998), the likelihood of committing Type I errors across all comparisons can be reduced by selecting a subset of both contaminant analytes and biological measures for pairwise comparisons.

As an example, a reduced list of chemicals was derived using the draft BERA COI screening tables³. All contaminants with average concentrations exceeding both the high and low NJDEP sediment benchmarks were identified and sorted by magnitude of the exceedance ratios (i.e., hazard quotients or HQs) (Attachment 2). Although the TCDD HQs are largest in both the freshwater only and overall LPR, this analyte can be ignored for the purposes of the SQT because it poses primarily a bioaccumulation hazard and is unlikely to contribute to biological responses observed in any of the SQT endpoints. Table 1 summarizes the remaining analytes (or aggregate sets of analytes) recommended for testing for the presence of significant concentration responses in the SQT analysis.

Similarly, one to two toxicity and benthic community endpoints (selected based on the specific area being evaluated and which demonstrates the best concentration responses) should be adequate.

¹ An important secondary question with both remedial and cost-allocation perspectives; however; additional (e.g., sediment Toxicity Identification Evaluations [TIE]) designed with this objective in mind would be necessary.

² Including five single species toxicity endpoints (i.e., *Ampelisca abdita* survival, *Chironomus dilutus* survival and biomass, and *Hyalella azteca* survival and biomass) and 6 benthic community metrics (i.e., abundance, taxa richness, Shannon-Weiner $-H'$, Pielou's J' , Swartz's Dominance Index, and Hilsenhoff Biotic Index).

³ LPRSA BERA App A_Attach A1_6-13-14.xlsx filtered on "sitewide sediment" and "sitewide sediment (>= RM 8)". NJDEP screening worksheets as presented in LPRSA BERA App A_Attach A2_6-13-14.xlsx.

Table 1. Summary of Recommended COI for Correlational Analysis

#	Freshwater	Entire
1	Total PAHs	Lead
2	Bis-(2-ethylhexyl)phthalate	Mercury
3	Phenol	Zinc
4	Total Chlordane	Total PAHs (or H/L PAHs)
5	Total DDx	Bis-(2-ethylhexyl)phthalate
6	Total PCB Congeners	Total DDx
7		Total PCB Congeners

References

Perneger, T.V., 1998. What's wrong with Bonferroni adjustments; *BMJ* 316:1236-1238.

Winward Environmental, LLC, 2009. Lower Passaic River Restoration Project. Lower Passaic River Study Area RI/FS. Quality Assurance Project Plan: Surface sediment chemical analyses and benthic invertebrate toxicity and bioaccumulation testing. Final. Prepared for Cooperating Parties Group, Newark, New Jersey. October 8, 2009. Winward Environmental LLC, Seattle, WA.

Winward Environmental, LLC, 2014. LPRSA Baseline Ecological Risk Assessment; June.

ATTACHMENT 1

Passaic River Analysis Framework for Biological Toxicity Metrics

Prepared by John Kern (KernStat) 8 September 2015

The Passaic River cooperating parties group (CPG) has conducted an evaluation of multiple contaminants of potential concern (COPCs) as they relate to ecological risks of adverse effects. This evaluation was composed of a series of bi-variate tests of hypothesis, the results of which were modified by Bonferroni correction to protect against potential False positives (i.e. Type I errors). Although the Bonferroni procedure has been widely applied, the procedure can lead researchers to miss potentially important biological effects. Failure to detect actual biological effects (i.e. false negatives) are more serious than false positive results because they may lead to ongoing unremediated risks to ecological resources.

The EPA recognizes that conducting many univariate tests can lead to false positive results simply by random chance and that adjusting test results to mitigate this potential may be necessary when many tests are conducted. An alternative to preserve the statistical power to detect real effects may be to conduct a smaller number of evaluations of more general questions followed. As an alternative to the large number of statistical hypothesis tests reported by CPG, EPA recommends that CPG conduct an alternative series of analyses that minimizes the number of hypothesis tests by evaluating broader, more general hypotheses using multiple generalized linear models. The proposed approach provides an integrated framework for evaluating the strength of evidence of biological effects while adjusting for potentially confounding covariates such as physical properties of the sediment, such as grain size, organic carbon content as well as characteristics of the organisms themselves such as lipid content.

The recommended approach progresses through an evaluation and quantification of potential biological effects, rather than deriving a simple yes or no answer as to the presence or absence of effects related to individual chemical compounds. A systematic framework of well -defined study questions is recommended in hopes of developing a clear evaluation of the role of chemical stressors related to ecological health of organisms in the sediment and water.

The overall goal of the evaluation is to develop a statistical evaluation of the likelihood of the following three general states;

- 1) Sediment chemistry is the primary stressor over and above the general effects of habitat quality and physical sediment characteristics,
- 2) Habitat quality and physical sediment characteristics are the primary stressors affecting ecological health and sediment chemistry is unimportant
- 3) A combination of chemistry, habitat and physical characteristics of the River are important predictors of ecological health, but the data are inadequate to separate their effects.

To develop statistical evidence supporting one or more of states one through three multiple variable analyses are recommended as a means to 1) acknowledge the potential that chemical compounds, physical sediment properties and organism specific characteristics may interact, 2) to minimize the overall number of statistical tests, thereby maximizing statistical power, and 3) to enhance precision of estimated effects by adjusting effects for covariation among stressors. With this approach, the potential that

biological effects are caused by mixtures of chemical compounds potentially covarying with physical sediment metrics is directly acknowledged and incorporated into the statistical analysis, providing a more complete description potential causative factors. Following is a description of an analysis framework the EPA recommends for the CPG to pursue as an alternative to the multiple testing with Bonferroni correction approach.

Methods

The questions described above anticipate that biological measurement endpoints may vary with the concentration of chemical stressors, the physical conditions of the sediment, variables specific to the biota itself or, most likely, a combination of all three. Multiple variable modeling methods, including multiple regression for normally distributed measurement endpoints or more generally multiple generalized linear models for non-normally distributed data are standard statistical tools used to estimate effects in the presence of confounding variables. The method known as analysis of covariance is one particular multiple regression model wherein measurement endpoints are compared among discrete populations, while adjusting for effects of a secondary continuous variable, such as percent lipid in the organism or organic carbon in sediment, or both. An extension to the ANCOVA model includes additional discrete or continuous variables, which may also confound or enhance effects of interest. Ideally a large multiple regression model could be constructed with each chemical concentration physical variable included as independent predictor variables. Tests for the importance of each chemical would be indicated by tests for the significance of regression coefficients on each predictor. However, it is recognized that some chemicals as well as physical variables may be correlated, so large multiple regression models including all of the chemicals of interest would violate the assumption that the predictors in a regression model are mutually independent. When predictors are correlated, regression coefficients may be biased and interpretation of results may be unreliable. In this situation the relative importance of individual predictors cannot be full identified, as measures of importance such as the partial coefficient of determination represent joint variation explained, rather than individual components of variation.

Recognizing that a large multiple regression model with many chemical and physical variables may not provide a reliable tool for assessing the importance of individual chemical stressors, EPA recommends that CPG utilize modern predictive modeling approaches to develop parsimonious models balancing model complexity (i.e. number of predictors) against model fit. Methods which are robust to correlation among predictors such as those described by Harrell (2001), are recommended. These approaches include reliance on model diagnostics, predictor variable reduction through multivariate methods such as principal components analysis and model testing using cross validation.

Of particular utility in this context may be the use of and multiple regression on the principal component scores of the analytical chemical data rather than directly on the chemistry metrics themselves. This approach is recommended because the principal component scores which represent weighted averages of the chemical concentrations are statistically independent and thereby satisfying the assumptions of multiple regression.

EPA recognizes that while multiple regression on principal component scores eliminates problems of multicollinearity among predictors, the resulting models may be more difficult to interpret. Principal component scores are weighted sums of the original data and as such generally represent mixtures of chemical compounds. Therefore, statistical inference based on this approach may be somewhat less satisfying because chemicals generally cannot be isolated individually. With this approach, fewer hypotheses are tested, assumptions of regression analysis are satisfied, but biological effects would be identified with general mixtures of chemical compounds, rather than with individual chemicals as might be preferred. However, it may be more realistic to expect that biological effects would be more closely tied to complex mixtures of chemicals that are present in the sediment.

Frequently the principal components represent groups of chemicals, such as organic vs inorganic compounds that tend to covary in the environment. For example the first principal component might be correlated with a combination of organic chemicals and other principal components might represent metals or pesticides. At times the unique principal components are not readily interpretable and in such a situation it may be informative to rotate the principal components to improve their interpretability. Seber (1977) terms this procedure Principal Factor analysis, although others simply refer to this as principal components analysis with rotation. Factor Analysis, a closely related multivariate method, known as factor analysis. Suhr (2003) provides a helpful review discussing the similarities and differences between these closely related methods, and Miesch (1980) provides additional discussion of the interrelationships between these methods and also discusses the importance or standardization of variables prior to extracting principal components. Because these methods are similar and frequently imprecisely referenced in the scientific literature, it is anticipated that specifics of selected computational methods and rationale for selection would be discussed in detail.

The EPA views this approach to statistical dimension reduction as a pragmatic alternative to testing many individual chemicals directly, with reduced power due to the need for Bonferroni adjustment. The alternative procedure based on principal component scores would compromise the potential to identify effects of specific individual chemical stressors while maintaining greater statistical power (because fewer tests would be conducted) to make more general statements about biological effects related to groups of chemical stressors. The number of tests is reduced because there would be only a single statistical test for each principal component entered into the regression as opposed to tests for all chemicals under consideration.

For continuous biological measurement endpoints the regression model would be of the form:

Biological Measurement Endpoint

$$= \beta_0 + \beta_1 \times PC_1 + \dots + \beta_k \times PC_k + \beta_{k+1}X_1 + \dots + \beta_{k+p}X_p + \varepsilon$$

Where the number of PC scores (k) would need to be selected, X_{k+i} represents to i^{th} physical variable of interest and ε represents a mean zero random error. The model would be varied somewhat for biological measurement endpoints that are either binary or counts using logistic or Poisson generalized linear models specific to these data types.

For this framework, regression coefficients for each PC score with their confidence intervals provide quantitative measures of the contribution of each group of chemicals to variation in the biological measurement endpoints. Additionally, because the physical variables are also in the simultaneous model, these coefficients represent the effect of the group of chemicals while controlling for differences in habitat.

As illustrated by Harrell (2001) multiple variable generalized linear model development is a broad topic with a range of potential approaches that could be implemented. The approach described above is one of several that could be considered as a means to develop a coherent framework for evaluating the effects of chemical stressors while controlling for covarying factors such as physical sediment characteristics, space and time. The EPA is open to other proposals from the CPG that are consistent with the general overall objective to estimate the effects of chemical stressors while statistically controlling for other potentially confounding environmental stressors. Other approaches could include more modern predictive modeling techniques such as classification and regression trees, machine learning approaches, or more traditional ordination techniques more familiar to ecologists. The EPA looks forward to working with the CPG to develop an analysis plan leading to a cogent description of the interrelationships between biological effects and environmental stressors.

References

- Harrell, F.E., Jr. 2001. *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis*. Springer Series in Statistics. Springer-Verlag, New York.
- Miesch, A.T. 1980. Scaling variables and interpretation of eigenvalues in principal components analysis of geological data. *Mathematical Geology*, Vol 12, No. 6. Plenum Publishing Company.
- Seber, G.A.F. 1977. *Multivariate Observations*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, New York.
- Suhr, D.D. 2003. Principal component analysis vs exploratory factor analysis. Proceedings of the 2003 SAS User Group Conference, Paper 2003-30. <http://www2.sas.com/proceedings/sugi30/203-30.pdf>; last visited 9/8/2015.

ATTACHMENT 2

Sediment Screening Evaluation

Table 2-1.
Summary of Sediment Benchmark Screening – Freshwater and Estuarine Habitats in
Lower Passaic River

Analyte Group	COPEC ¹	NJDEP Sediment Benchmark Exceedances ²		Notes
		Lower-bound	Upper-bound	
Freshwater				
PCDDs/PCDFs	2,3,7,8-TCDD	6.93E+01	7.18E+02	bioaccumulation hazard primarily ³
SVOCs	Bis-(2-ethylhexyl)phthalate	1.47E+01	6.04E+01	
PAHs	Chrysene	9.13E+00	1.24E+01	Total PAHs (or H/L PAHs)
PCB Aroclors	Aroclor-1254	9.12E+00	5.17E+00	Total PCBs
OC Pesticides	Total Chlordane	8.83E+00	1.64E+01	
OC Pesticides	Total DDx	8.33E+00	1.89E+01	
PAHs	Pyrene	7.06E+00	1.22E+01	Total PAHs (or H/L PAHs)
PAHs	Benzo(g,h,i)perylene	6.88E+00	1.29E+01	Total PAHs (or H/L PAHs)
PAHs	Fluoranthene	6.47E+00	8.80E+00	Total PAHs (or H/L PAHs)
PAHs	Indeno(1,2,3-cd)pyrene	6.25E+00	1.00E+01	Total PAHs (or H/L PAHs)
PCB Aroclors	Aroclor-1260	6.25E+00	3.00E+01	Total PCBs
OC Pesticides	4,4'-DDD	5.50E+00	4.13E+00	Total DDx
SVOCs	Phenol	5.00E+00	4.90E+00	
PAHs	Phenanthrene	4.63E+00	7.86E+00	Total PAHs (or H/L PAHs)
PAHs	Total PAHs	4.10E+00	2.55E+01	
PAHs	Fluorene	3.69E+00	3.11E+00	Total PAHs (or H/L PAHs)
PAHs	Dibenzo(a,h)anthracene	3.23E+00	7.00E+00	Total PAHs (or H/L PAHs)
PAHs	Anthracene	2.70E+00	4.55E+00	Total PAHs (or H/L PAHs)
PCB Congeners	Total PCB Congeners	2.45E+00	2.17E+01	
PAHs	Benzo(a)pyrene	2.43E+00	9.46E+00	Total PAHs (or H/L PAHs)
PAHs	Benzo(a)anthracene	2.09E+00	9.69E+00	Total PAHs (or H/L PAHs)
OC Pesticides	4,4'-DDE	1.95E+00	7.40E+00	Total DDx
PAHs	Benzo(k)fluoranthene	1.49E+00	8.33E+00	Total PAHs (or H/L PAHs)
PAHs	Acenaphthene	1.08E+00	3.38E+01	Total PAHs (or H/L PAHs)
Entire Lower 17-Miles				
PCDDs/PCDFs	2,3,7,8-TCDD	3.33E+02	4.80E+02	bioaccumulation hazard primarily ³
PCB Congeners	Total PCB Congeners	8.89E+00	7.05E+01	
SVOCs	Bis-(2-ethylhexyl)phthalate	5.77E+00	8.33E+01	
OC Pesticides	4,4'-DDT	4.71E+00	3.30E+01	Total DDx
Metals	Mercury	3.38E+00	6.49E+01	
OC Pesticides	Total DDx	3.26E+00	9.49E+01	
PAHs	Benzo(a)pyrene	2.00E+00	7.44E+00	Total PAHs (or H/L PAHs)
OC Pesticides	4,4'-DDE	1.96E+00	2.41E+01	Total DDx
OC Pesticides	4,4'-DDD	1.85E+00	1.85E+01	Total DDx
PAHs	Phenanthrene	1.73E+00	1.08E+01	Total PAHs (or H/L PAHs)
PAHs	Pyrene	1.69E+00	6.62E+00	Total PAHs (or H/L PAHs)
PAHs	Benzo(a)anthracene	1.63E+00	9.96E+00	Total PAHs (or H/L PAHs)
PAHs	Dibenzo(a,h)anthracene	1.58E+00	6.51E+00	Total PAHs (or H/L PAHs)
PAHs	Chrysene	1.29E+00	9.38E+00	Total PAHs (or H/L PAHs)
Metals	Zinc	1.12E+00	3.07E+00	
Metals	Lead	1.10E+00	2.26E+01	

Footnotes:

1. Analytes from COPEC screening tables (Windward, 2014) included if hazard quotients for both lower- and upper-bound NJDEP sediment benchmarks exceed 1.
2. Ratio of maximum sediment concentration divided by the lower- or upper-bound sediment threshold concentration.
3. Although site-specific reproductive effects in oysters have been documented, a concentration response between TCDD and SQT measurement endpoints is unlikely.